VIROLOGY 206, 718–723 (1995)

Rabies Virus Glycoprotein Gene Contains a Long 3′ Noncoding Region Which Lacks Pseudogene Properties[1]

EUGENE V. RAVKOV,*,† JEAN S. SMITH,† AND STUART T. NICHOL*,†,[2]

*Cell and Molecular Biology Program and Department of Biochemistry, University of Nevada, Reno, Nevada 89557; and †Division of Viral and Rickettsial Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia 30333

Received July 19, 1994; accepted September 26, 1994

Analysis of a limited number of laboratory strains of rabies virus had demonstrated the presence of a genome region bounded by two transcription termination and polyadenylation-like (TTP) signals (approximately 400 to 450 nucleotides apart) which was located between the end of the glycoprotein (G) coding sequence and the beginning of the L polymerase coding sequence. Although this region had been suggested to represent a remnant or pseudogene (ψ), no detailed analysis had been carried out to examine this possibility. Here we present the nucleotide sequence analysis of this genome region for several laboratory rabies virus strains and a large number of diverse rabies viruses detected directly in brain tissue of naturally infected animals. Only one distinct lineage of the laboratory strains and none of the wild-type rabies viruses contained the upstream TTP-like signal, indicating that only the downstream TTP motif is the authentic G mRNA transcription termination and polyadenylation and signal. Phylogenetic analysis of sequence differences provided no evidence of laboratory strains containing the two TTP-like signals being ancestral to any of the viruses possessing only the downstream TTP sequence motif. These data indicate that this region of the rabies virus genome encodes a G mRNA with a long 3′ noncoding region with no evidence of a pseudogene.

Rabies virus is a member of the *Lyssavirus* genus of the *Rhabdoviridae* family, a diverse group of animal, plant, and insect viruses, including vesicular stomatitis virus (VSV), rabies and related viruses, infectious hematopoietic necrosis virus (IHNV), and bovine ephemeral fever virus (BEFV) (*1, 2*). These viruses all possess nonsegmented, negative-sense, single-stranded RNA genomes of approximately 10 to 15 kb in length. The virus genomes commonly encode at least five mRNAs, which are located and also sequentially transcribed in the general pattern 3′-nucleocapsid protein (N), polymerase-associated phosphoprotein (P or M1), matrix protein (M or M2), glycoprotein(s) (G), and large polymerase protein (L)-5′. These genomes usually contain very little noncoding nucleotide sequence and gene junctions are small. The most common exception to this overall genetic structure is the genome region located between the G mRNA termination codon and the L mRNA initiation signal. In IHNV, an additional gene encoding a nonstructural protein (NV) is located in between the G and L genes (*3*). BEFV contains an additional nonstructural glycoprotein (G_{NS}) in this genome area that has been proposed to have evolved by duplication of the original glycoprotein gene (*4*). Further, other small open reading frames (ORFs), α and β, are located between G_{NS} and the beginning of the L gene. Similarly, rabies virus was found to possess a long (approximately 425 nucleotide) region of noncoding sequence between the G and L genes (*5*) that was suggested to represent a remnant or pseudogene (ψ) which was once functional, but had become vestigial.

Two transcription termination and polyadenylation (TTP) motifs were found in this region in the Pasteur strain (PAS PV) of rabies derived from an 1882 French rabies isolate (*5*). The first was 70 nucleotides downstream from the translation stop codon of the G gene and the other was located at the end of the proposed ψ region, 24 nucleotides upstream of the L gene start. This arrangement suggested that the upstream motif was the TTP signal for G mRNA transcription and the downstream motif was presumably related to the previous function of the pseudogene. The presence of both upstream and downstream TTP sites in a second rabies strain (SAD, derived from a 1935 U.S. rabies isolate) suggested that this may be a common gene arrangement in rabies viruses (*6*). The complexity of this issue became apparent when it was shown that two species of G mRNA (1.9 and 2.3 kb in length) were synthesized in cells infected with the ERA rabies strain (*7*). The shorter mRNA was generated by use of the upstream TTP, and the larger mRNA by the use of the downstream TTP. In addition, another rabies laboratory strain, HEP (derived from a 1939 U.S. rabies isolate), lacked the upstream TTP and synthesized

[1] Use of trade names and commercial sources is for identification only and does not imply endorsement by the Public Health Service or the U.S. Department of Health and Human Services.

[2] To whom correspondence and reprint requests should be addressed. Fax: (404)639-1118. E-mail: stn1@ciddvd1.em.cdc.gov.

0042-6822/95 $6.00

only the 2.3-kb mRNA by use of the downstream TTP. Later studies of two additional virus strains reportedly derived from the Pasteur virus (CVS and Nishigahara) also found only the downstream TTP motif (*8, 9*) raising the question as to which of the TTP signals was ubiquitous in the synthesis of the G mRNA in wild-type rabies virus isolates. It also cast some doubt on considering the long noncoding region between G and L as a pseudogene.

In order to gain a better understanding of the origin and potential function of this region of the rabies virus genome and to investigate apparent discrepancies between actual and recorded passage histories of several common rabies virus laboratory strains, a genetic analysis of the region was initiated. The basic approach involved the PCR amplification and nucleotide sequence analysis of this genome region for a broad selection of street rabies infected material and comparison with nucleotide sequences of the equivalent genome region of numerous laboratory strains.

Virus field samples to be analyzed were selected to represent a spectrum of the animal species which serve as reservoirs for rabies (*10, 11*). The geographic area and host species defining these reservoirs and the sample source and year of collection are as follows: USTXCY1386 coyote, 1991, central Texas [represents virus reservoir in the gray fox (*Urocyon cinereoargenteus*)] USTXDG1259 domestic dog (*Canis familiaris*), 1991, border area of Texas–Mexico; USARSK963 striped skunk (*Mephitis mephitis*), 1984, northern Arkansas; USARSK964 striped skunk (*M. mephitis*), 1984, southern Arkansas; USFLRC623 raccoon (*Procyon lotor*), 1987, Florida; USFLBT884 hoary bat (*Lasiurius cinereus*), 1984, Florida; USFLBT856 Brazilian freetail bat (*Tadarida brasilienses*), 1988, Florida; ARBADG116 domestic dog, 1979, Buenos Aires, Argentina; CNONFX5 (type 5) red fox (*Vulpes vulpes*), 1990–1992, Ontario; and FRFXWR56 red fox, 1990, France. All samples were brain material obtained directly from naturally infected animals, with the exception of sample ARBADG116, which had been serially passaged in mice and cell culture. Nucleotide sequence for sample CNONFX5 (*12*) was obtained from GenBank (Accession No. U03765). Nucleotide sequence for sample FRFXWR6 was published previously (*8*).

Several vaccine and laboratory strains were also included in the analysis. The histories of these strains have been published previously (*8, 9, 13, 14,* and H. N. Johnson, Rabies Information Exchange No. 12, June 1985). Briefly, these included: (1) the PAS Pasteur virus was originally isolated from a cow infected by a dog in a Paris suburb in 1882 and serially passaged in rabbits and cell culture at Pasteur Institute, Paris; (2) PAS PV, a second stock of Pasteur virus, obtained in 1965 from the Pan-American Zoonoses Center (Buenos Aires, Argentina); (3) CVS (challenge virus standard), used for rabies vaccine potency tests, was selected in 1940 by the National Institutes of Health (U.S.A.) from greater than 30 stocks of Pasteur virus in different laboratories in the United States and distributed worldwide (sample was virus stock at the Centers for Disease Control and Prevention); (4) NISHIGAHARA virus derived from stock of Pasteur virus received before 1915 at National Institutes of Health of Domestic Animals, Nishigahara, Tokyo, and serially passaged in rabbits, guinea pigs, embryonated eggs, and cell culture; (5–7) SAD/ERA, Street Alabama (SA) isolated in 1935 from a dog in Alabama (U.S.A.), the virus stock was serially passaged in mice and distributed in the 1940s by the National Institutes of Health (U.S.A.) as a vaccine challenge to represent a virus of North American origin compared with the European origin of the Pasteur virus. The virus stock held by the Dufferin Division of Connaught Laboratories Ltd. (SA-D) was attenuated by serial passage in cell culture and was the source of the SADB19 and ERA vaccine strains; (5) SAD1520 was a stock of SAD which is presently used as vaccine seed by a U.S. pharmaceutical company; (6) SAD B19 virus derived from SAD and attenuated by multiple passages in cell culture; (7) ERA virus derived from SAD and attenuated by multiple passages in embryonated eggs and cell culture (name derived from the investigators, Evelyn–Rokitnicki–Abelseth); (8) LEP (low egg passage) virus isolated in 1939 from a human infected by a dog in Georgia (U.S.A.) and attenuated by passage in embryonated eggs (LEP was the parent virus for high egg passage (HEP) virus). SAD1520, LEP, and CVS were sequenced in this study. Additional nucleotide sequences for vaccine and laboratory strains were obtained from GenBank (Accession Nos. M13215, M31046, and J02293) and from published data (*8, 9*).

The nucleotide sequence of an approximately 440-nucleotide region of the genome of the viruses was determined from around 40–50 nucleotides downstream of the translation stop codon of the G mRNA to 20–24 nucleotides upstream of the start codon of the L mRNA. The nucleotide sequence differences of these viruses relative to the PAS and PAS PV standard strains are shown in Fig. 1. The entire aligned sequences shown are 528 nucleotides in length. Most of the sequences derived in this study are shown from aligned position 70–90 to 528. The sequence of the CVS strain reported here is virtually identical with that reported earlier for a different stock of the virus (*8*). A high degree variability was found, with up to 42.8% nucleotide sequence difference observed (for bases 79 to 528) among the virus isolates. Several regions, particularly bases 75 to 125 (which encompasses the upstream potential TTP area) and 290 to 410, are highly diverse, although distinct regions of sequence conservation (e.g., nucleotide bases 125 to 280 and 400 to 528) can also be seen (Fig. 1). In addition, several relative insertions or deletions are evident between the various virus sequences. Analysis of the entire sequence region (using FRAMES and MFOLD

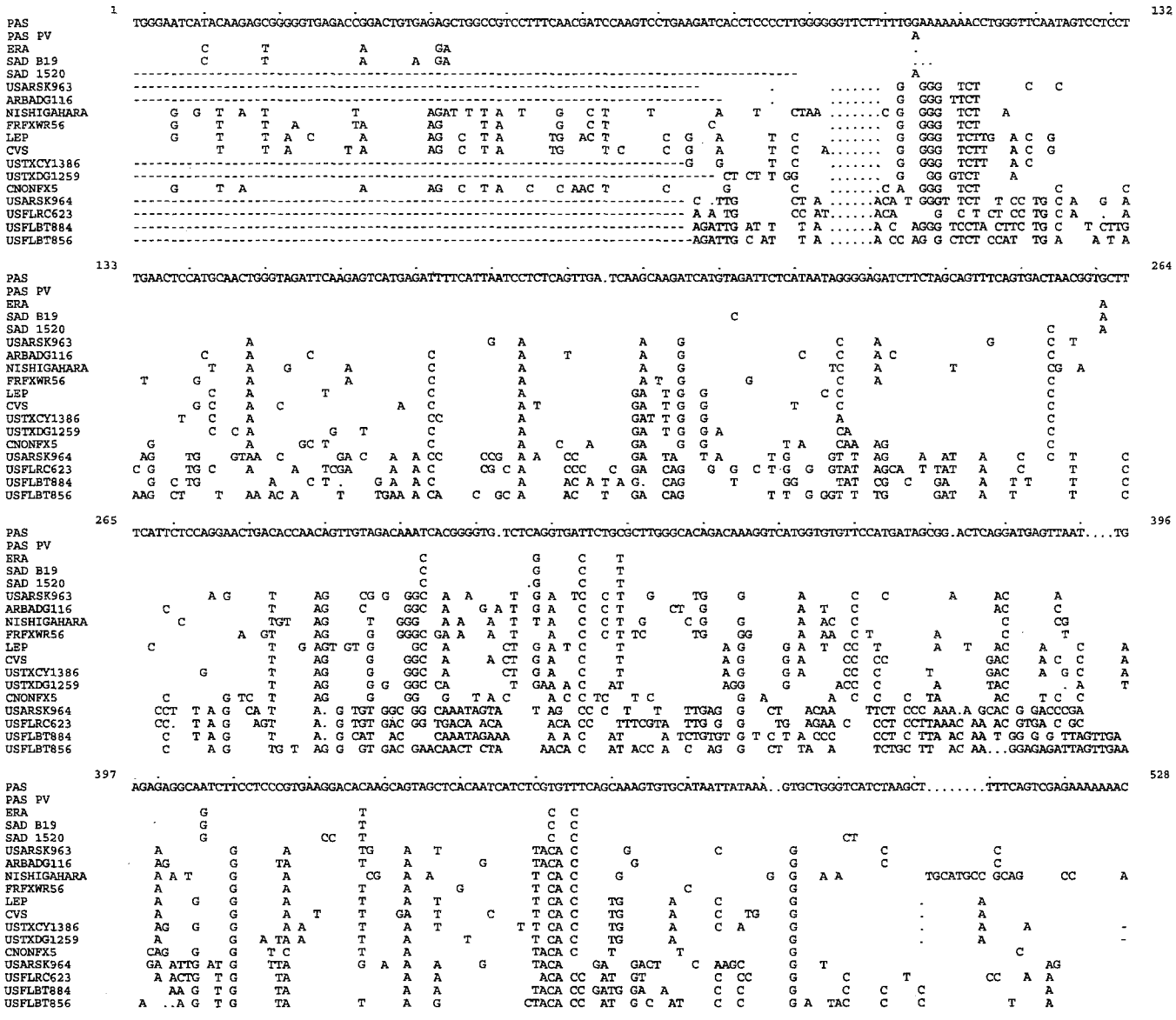

FIG. 1. Genetic variability of the G–L noncoding region of rabies virus laboratory and wild-type strains. Extraction of total RNA from brain material and infected cells and polymerase chain reaction (PCR) amplification of the virus G–L RNA region were carried out as described previously (*15–17*). Primers G(+) (position 4665 to 4687) and L(−) (position 5543 to 5520) were utilized as reported earlier (*8*). All primers are numbered relative to the PV strain genome sequence (*5*; GenBank Accession No. M132115). Purified product DNA (approximately 850 bp in size) was sequenced either directly by the dideoxy-nucleotide method or following cloning into pCRII vector using the TA cloning kit (InVitrogen, La Jolla, CA) depending on the intensity of the PCR product band. Primers Gnc2(−) (5129 to 5149), Gnc1(−) (5096 to 5119), and Gnc(+) (5085 to 5106) were utilized in sequence analysis. When nucleotide sequences were obtained from DNA clones, multiple clones were analyzed to verify the correct consensus sequence with the exception of samples USFLBT856 and USFLBT884, for which only one clone was obtained. Nucleotide sequence analysis and alignment used the PILEUP and LINEUP programs of the GCG Sequence Analysis Software Package, Version 7 (GCG, Madison, WI). Sequences are shown in the cDNA (i.e., positive) sense and differences relative to the PAS strain are indicated. Dashes indicate undetermined sequence regions. Dots indicate gaps inserted to optimize sequence alignments.

programs of the GCG software package) failed to reveal evidence of conservation of any ORFs of significant length or RNA secondary structural features within this region (data not shown).

Both upstream and downstream potential TTP signals can be seen in the five samples of PAS- and SAD-derived viruses which are very closely related, displaying less than 3.3% nucleotide sequence divergence (Fig. 2). None of the other laboratory strains (including two strains reportedly derived from PAS) and none of the diverse wild-type strains (displaying between 14.2 to 42.8% divergence) were found to contain the potential upstream TTP signal at the expected position. Only the downstream TTP signal, the function of which has been suggested to be associated with the remnant pseudogene, was conserved in all the viruses analyzed. Comparison of the

Virus	79	127	510	528
PAS	UGGAGGGGAACCCCCCAAGAAAAACCUUUUUUUGGACCCAAGUUAUCAG		AAAGUCAGCUCUUUUUUUG	
PAS PV	UGGAGGGGAACCCCCCAAGAAAAACUUUUUUUUGGACCCAAGUUAUCAG		AAAGUCAGCUCUUUUUUUG	
ERA	UGGAGGGGAACCCCCCAAGAAAAAC.UUUUUUUGGACCCAAGUUAUCAG		AAAGUCAGCUCUUUUUUUG	
SAD B19	UGGAGGGGAACCCCCCAAGAAAAAC...UUUUUGGACCCAAGUUAUCAG		AAAGUCAGCUCUUUUUUUG	
SAD 1520	----------CCCCCCAAGAAAAACUUUUUUUUGGACCCAAGUUAUCAG		AAAGUCAGCUCUUUUUUUG	
USARSK963	UGGAGG.GAACCCC.....AACCCCCCUUAGAGACCCAGGUUGUCAG		AGAGUCAGCUCUUUUUUUG	
ARBADG116	UGGAGG.GAACCCC.....AACCCCCCUAAGAGACCCAAGUUAUCAG		AGAGUCAGCUCUUUUUUUG	
NISHIGAHARA	UGGAAGGGAGAUUC.....GACCCCCCUUAGAGACUCAAGUUAUCAG		ACGUCCAGCUGGUUUUUUU	
FRFXWR56	UGGAGGGGAACCCC.....AACCCCCCUUAGAGACCCAAGUUAUCAG		AAAGUCAGCUCUUUUUUUG	
LEP	UGGAGGAGAGCCCC.....AACCCCCCUUAGAACCCUAGGUCAUCAG		AAAGUCAGCUCUUUUUUUG	
CVS	UGGAGGAGAGCCCU.....AACCCCCCUUAGAAACCUAGGUCAUCAG		AAAGUCAGCUCUUUUUUUG	
USTXCY1386	UGGAGGAGAGCCCC.....AACCCCCCUUAGAAACCUAGGUUAUCAG		AAAGUUAGCUCUUUUUUU-	
USTXDG1259	GAGGAGAGCCCCCC.....AACCCUCCUCAGAGACCUAAGUUAUCAG		AAAGUCAGCUCUUUUUUU-	
CNONFX5	CGGAGGGGAGCCCC.....GAUCCCCCUUAGAGACCCAAGUUGUCAG		AAAGGCAGCUCUUUUUUUG	
USARSK964	ACGAGGGGAGACUC.....UGUAACCCCAUAGAGAACGGAACUGUUAG		AAAGUCAGUCCUUUUUUUG	
USFLRC623	ACGAGGGGAGGCUA.....UGUACCUUCUUGUAGGACGGAACUGUUAG		GGAGUUAGUUCUUUUUUUG	
USFLBT884	ACGUAGAGAAACUC.....UAGACUCCCUAGGAUAGAAGAACUGUCAA		AAAGUCAGUUCUUUUUUUG	
USFLBT856	ACGGGUAGAAACUC.....UAGGCUCUCUGAGAGGGUAAAACUUUCAG		AAAAUCAGUUCUUUUUUUG	

FIG. 2. Conservation and variation of TTP-like motifs located between G and L gene coding sequences. Sequences are shown in the RNA genome (i.e., negative sense). Both upstream and downstream U-rich regions are underlined. Dashes indicate undetermined sequence regions. Dots indicate gaps inserted to optimize sequence alignments.

conservation of the upstream and downstream TTP signals shows that even among the five laboratory strains which do possess the upstream TTP signal, this signal is not as highly conserved as the downstream TTP. The number of uridine residues in the upstream TTP site ranges from 5 to 8 nucleotides, in contrast to the downstream G mRNA TTP site, where the seven uridine residues of this signal motif are highly conserved in all the sequences analyzed.

A phylogenetic analysis of nucleotide sequence differences (bases 79 to 528) was carried out to determine if the five viruses possessing both TTP signals were estimated to be ancestral to those that contained only the downstream TTP signal. Two equally parsimonious trees were obtained. One is shown in Fig. 3, the other had the minor variation of the Nishigahara and FRFXWR56 samples not being in the same clade as USARSK963 and ARBADG116 but representing a separate closely related clade. The five samples of PAS- and SAD-derived viruses containing both TTP signals were found to share a high degree of sequence identity and form a genetic lineage distant from the predicted ancestral node of the tree. This is inconsistent with these viruses having an ancestral relationship to the other diverse virus lineages which possess only the downstream TTP signal. As expected the wild-type viruses which had been chosen as representatives of various virus types were distributed throughout the tree in the form of multiple distinct lineages. A similar phylogeny is also obtained on analysis of N gene nucleotide sequences (*11, 16, 25* and J. S. Smith, unpublished).

Since the original suggestion of a pseudogene by Tordo and colleagues (5), several subsequent rabies articles have labeled the long noncoding sequence region located between the rabies G gene stop codon and the L gene initiation codon as the remnant or pseudogene (ψ) region (*6, 8, 20–23*). The primary supporting evidence for a pseudogene is the presence of 5 to 8 U residues representing a potential TTP site, located approximately 50 nucleotides downstream of the translation termination codon of the G mRNA in the closely related PAS and

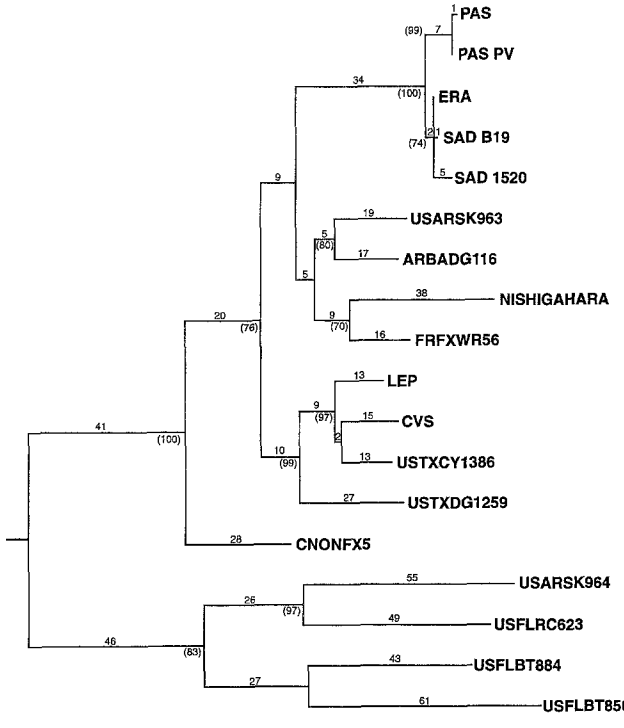

FIG. 3. Phylogenetic relationship between rabies viruses based on nucleotide sequence differences of noncoding region between G and L genes. Phylogenetic analysis of nucleotide sequence differences was done by the maximum parsimony method using the PAUP software program (*18*) run on a Macintosh Quadra 800 (Apple). Analysis used the MULPARS and branch and bound options. The equivalent genome region of Mokola virus (*19*) was included to outgroup root the analysis. Bootstrap confidence limits were calculated by 1000 repetitions of the analysis and limits in excess of 50% are indicated in parentheses at branch points. Horizontal lengths are proportional to nucleotide step differences (indicated adjacent to line) and vertical distances are for graphic presentation only.

ERA strains. Also, a very high level of genetic stability was anticipated for all rabies isolates based on the close genetic relatedness of these two viruses having recorded histories of isolation more than 50 years apart and on different continents. However, it now appears that the recorded histories of the first two virus strains to be sequenced are incorrect. The molecular phylogenies of these and other laboratory strains of rabies suggest that several strains were misidentified, most likely in the United States prior to 1940, and that some conclusions drawn from the early studies are incorrect.

During the early part of this century, aliquots of the Pasteur vaccine virus were distributed to vaccine production facilities and public health laboratories worldwide. Low-temperature storage was not widely available, and frequent serial passage of the virus stocks was necessary to maintain viability. By 1940, marked differences in the antigenicity and virulence of stocks of Pasteur virus held in different laboratories in the United States were noted (*24*). The National Institutes of Health, in collaboration with the Rockefeller Foundation and public health laboratories in Alabama, Georgia, and New York organized comparative potency tests of selected reference stocks of vaccine seed and challenge virus standard viruses. With the exception of the LEP/HEP virus, all the laboratory and vaccine strains presently in use in the United States are derived from stocks of virus distributed in these studies (H. N. Johnson, Rabies Information Exchange No. 12, June 1985). Given the incomplete and often inconsistent histories for many of the laboratory viruses, it may not be possible to identify precisely the true origin of any of these viruses. The history of the Nishigahara virus strain is well documented as derived from a Pasteur virus stock sent to Japan prior to 1915. Although this virus is found in a different area of the phylogenetic tree from the PAS- and SAD-related viruses, only Nishigahara virus, of all the laboratory strains recorded to have been derived from the Pasteur virus, is found to be genetically closely related and monophyletic with contemporary rabies viruses found in France. The simplest interpretation is that Nishigahara may best represent virus strains of the original Pasteur type. However, these results should be interpreted with care, as the introduction of rabid dogs from Europe to Asia, Africa, and North and South America during 18th and 19th century colonization efforts is well documented and evidence is accumulating for the worldwide occurrence of rabies virus strains originally of European lineage (*16, 25*).

Irrespective of their recorded histories, it is becoming clear that the viruses currently identified as PAS, PAS-PV, SAD, and ERA are multiple samples of essentially the same virus and that the first TTP region is present only in this virus stock. In sharp contrast, the downstream TTP site (450 nucleotides further downstream, and only 25 nucleotides upstream of the L gene initiation site) is extremely conserved in all viruses analyzed. With the exception of one laboratory strain, Nishigahara, the motif CUUUUUUUG is found in all viruses for which the complete sequence of this area is available. Even in Nishigahara, the central seven U residue motif is conserved.

It appears counterintuitive that the signal suggested to be associated with the functional G mRNA is poorly conserved, but the signal suggested to be associated with a remnant pseudogene is highly conserved. In spite of this, one could speculate that the viruses containing both TTP sites are representatives of the ancestral form of rabies viruses and that the other strains have lost this signal. However, the phylogenetic analysis shows that this is not the case; the strains containing both potential TTP sites are all closely related and form one distinct lineage which does not appear ancestral to the other multiple rabies lineages. In addition, careful analysis of the coding potential of this genome region failed to reveal any hint of an underlying remnant coding frame. Taken together, these data provide no support for this long noncoding region representing a pseudogene. The accumulated evidence indicates that this region is the G mRNA 3′ noncoding region. Long 3′ noncoding regions have been found in other negative strand RNA viruses (e.g., morbilliviruses, *26*; filoviruses, *27;* sonchus yellow net virus, *28*), including the G mRNA of another rhabdovirus, VSV (Indiana serotype) from Central America (*29*). The exact function of these noncoding sequences is still unknown, but they may influence mRNA interaction with host proteins and RNA stability. In several other negative-strand RNA viruses, replication errors such as polymerase skipping or slippage events generating repeated sequences are thought to have been involved in the generation of long RNA 3′ noncoding sequences (*29–31*). Some evidence of imprecise repeats can be seen throughout the rabies G mRNA 3′ noncoding sequence (data not shown). Similarly, the upstream U-rich motif which appears to have been gained by the PAS- and SAD-related strains may well have been acquired by polymerase stuttering, thus generating untemplated U residues at this position. The incorporation of U-rich sequences by polymerase stuttering is thought to have been involved in the generation of long G mRNA 3′ noncoding sequences in some VSV Indiana serotype viruses (*29*). It is interesting that runs of single nucleotides appear around this rabies genome area, including stretches of C and G residues (Fig. 2). The G + C-rich nature may increase local secondary structure and make transit by the virus polymerase more difficult, possibly promoting potential slippage events in this vicinity.

The possible role of an additional TTP site in the G gene in the alteration of virulence of the PAS and SAD type virus strains is difficult to evaluate. It can be speculated that as a proportion of active polymerase molecules detach from the virus transcribing ribonucleoprotein complex at each TTP site (*32*), an additional TTP near the end of the G gene could cause a down regulation of

the amount of the downstream L mRNA synthesized. In addition, polymerase molecules terminating G mRNA synthesis at the upstream TTP would have to transit an extra approximately 400 nucleotide long intragenic region before initiating L mRNA synthesis. This may increase the likelihood of polymerase detachment between termination of G and initiation of L mRNA synthesis. Finally, shortening of G mRNA 3′ noncoding sequences by over 400 nucleotides may alter mRNA stability and down regulate G protein synthesis. However, such events are clearly not involved in the attenuation of the HEP/LEP strain, which lacks the additional TTP site.

Prior to the development of PCR, the large amounts of RNA required for nucleotide sequencing limited genetic studies to analysis of laboratory-adapted strains of rabies. The data presented here and by others (*7–9*) reveal more genetic variation among the rabies virus isolates from different reservoirs and suggest that the phylogeny of rabies virus is dominated more by geographic and host origin factors than by temporal factors, which probably reflects the strong host selective pressures acting during the maintenance and transmission of the virus in nature. Genetic analysis of a larger number of rabies viruses and samples from diverse geographic regions and host species will shed further light on the exact origins of some of the classic laboratory strains and provide important information on the natural history and epidemiology of this ancient disease.

ACKNOWLEDGMENTS

We thank Michael Bowen, Tony Sanchez, and Heinz Feldmann for assistance and critical comments on the manuscript. We also thank John O'Connor for his editorial assistance. This work was supported by U.S.D.A. Animal Molecular Biology National Research Initiative Grant 90-37266-5473.

REFERENCES

1. WUNNER, W. H., *In* "The Rhabdoviruses" (R. R. Wagner, Ed.), pp. 361–426. Plenum, NY, 1987.
2. WUNNER, W. H., CALISHER, C. H., DIETZGEN, R. G., JACKSON, A. O., KITAJIMA, E. W., LAFON, M. F., LEONG, J. C., NICHOL, S. T., PETERS, D., SMITH, J. S., and WALKER, P. J., *In* "Classification and Nomenclature of Viruses" Springer-Verlag, NY, 1994, in press.
3. KURATH, G., and LEONG, J., *J. Virol.* **53,** 462–468 (1985).
4. WALKER, P. J., BYRNE, K. A., RIDING, G. A., COWLY, J. A., WANG, Y., and MCWILLIAM, S., *Virology* **191,** 49–61 (1992).
5. TORDO, N., POCH, A., ERMINE, A., KEITH, G., and ROUGEON, F., *Proc. Natl. Acad. Sci. USA* **83,** 3914–3918 (1986).
6. CONZELMANN, K. K., COX, J. H., SCHNEIDER, L. G., and THIEL, H. J., *Virology* **175,** 485–499 (1990).
7. MORIMOTO, K., OHKUBO, A., and KAWAI, A., *Virology* **173,** 465–477 (1989).
8. SACRAMENTO, D., BADRAN, H., BOURHY, H., and TORDO, N., *J. Gen. Virol.* **73,** 1149–1158 (1992).
9. SAKAMOTO, S. I., IDE, T., NAKATAKE, H., TOKIYOSHI, S., YAMAMOTO, M., KAWAI, A., and SMITH, J., *Virus Genes* **8,** 35–46 (1994).
10. SMITH, J. S., *Adv. Virus Res.* **36,** 215–253 (1989).
11. RUPPRECHT, C. E., and SMITH, J. S., *Semin. Virol.* **5,** 155–164 (1994).
12. NADIN-DAVIS, S. A., CASEY, G. A., and WANDELER, A., *J. Gen. Virol.* **74,** 829–837 (1993).
13. PASTEUR, L., *Semaine Médicale* **4,** 318–320 (1884).
14. CLARK, H. F., and WIKTOR, T. J., *In* "Strains of Human Viruses" (M. Majer and S. A. Plotkin, Eds.), pp. 177–182. Karger, Basel, Switzerland, .1972.
15. SMITH, J. S., FISHBEIN, D. B., RUPPRECHT, C. E., and CLARK, K., *N. Engl. J. Med.* **324,** 205–211 (1991).
16. SMITH, J. S., ORCIARI, L. A., YAGER, P. A., SEIDEL, H. D., and WARNER, C. K., *J. Infect. Dis.* **166,** 296–307 (1992).
17. NICHOL, S. T., ROWE, J. E., and FITCH, W. M., *Proc. Natl. Acad. Sci. USA* **90,** 10411–10420 (1993).
18. SWOFFORD, D. L., "PAUP: Phylogenetic Analysis Using Parsimony," version 3.0s. Computer program. Illinois Natural History Survey, Champaign, IL, 1991.
19. BOURHY, H., KISSI, B., and TORDO, N., *Virology* **194,** 70–81 (1993).
20. TORDO, N., POCH, A., ERMINE, A., KEITH, G., and ROUGEON, F., *Virology* **165,** 565–576 (1988).
21. CONZELMANN, K. K., COX, J. H., and THIEL, H. J., *Virology* **184,** 655–663 (1991).
22. SACRAMENTO, D., BOURHY, H., and TORDO, N., *Mol. Cell. Probes* **5,** 229–240 (1991).
23. BAER, G. M., and TORDO, N., *In* "Encyclopedia of Virology" (R. G. Webster and A. Granoff, Eds.), pp. 1180–1185. Academic Press, San Diego, CA, 1994.
24. HABEL, K., *Publ. Health Rep.* **55,** 1619–1634 (1940).
25. SMITH, J. S., YAGER, P. A., and ORCIARI, L. A., *Onderstepoort J. Vet. Sci.* **60,** 307–314 (1994).
26. BARON, M. D., GOATLEY, L., and BARRETT, T., *Virology* **200,** 121–129 (1994).
27. SANCHEZ, A., KILEY, M. P., HOLLOWAY, B. P., and AUPERIN, D. D., *Virus Res.* **29,** 215–240 (1993).
28. HEATON, L. A., HILLMAN, B. I., HUNTER, B. G., ZUIDEMA, D., and JACKSON, A. O., *Proc. Natl. Acad. Sci. USA* **86,** 8665–8668 (1989).
29. BILSEL, P. A., and NICHOL, S. T., *J. Virol.* **64,** 4873–4883 (1990).
30. DUNN, E. F., PRITLOVE, D. C., and ELLIOT, R. M., *J. Gen. Virol.* **75,** 597–608 (1994).
31. SPIROPOULOU, C. F., MORZUNOV, S., FELDMANN, H., SANCHEZ, A., PETERS, C. J., and NICHOL, S. T., *Virology* **200,** 715–723 (1994).
32. IVERSON, L. E., and ROSE, J. K., *Cell* **23,** 477–484 (1981).